A phenotypically driven segmentation for 3D facial morphology

DZEMILA SERO[1], MARK SHRIVER[2], DIRK VANDERMEULEN[1], PETER CLAES[1]

[1]KU Leuven, ESAT/PSI - UZ Leuven, MIRC

[2]Department of Anthropology, Pennsylvania State University, University Park, PA, USA

Corresponding author: Dzemila Sero, dzemila.sero@kuleuven.be, +32 16 34 90 32

Abbreviated title: Modularity of 3D faces through Spectral Clustering

ABSTRACT

Facial morphology is the result of mazy interactions between environmental and epigenetic factors that lead to the composition of multiple subunits integrated to function as a whole. In this work, we combine modularity concepts from evolutionary developmental biology with unsupervised machine learning tools to provide a descriptive framework of the facial configuration of landmarks on a modular basis. We apply normalized spectral clustering to a database of 592 3D faces - represented with spatially dense meshes of 7,150 quasi-landmarks -, grouping vertices that are strongly correlated and connected to form compact modules. We first build the affinity matrix that encodes the structural similarity, both in terms of correlation and distance, between each pair of landmarks. The normalized spectral clustering is then applied on the affinity matrix built as such. Since the strength of co-variation between the obtained modules is the criterion for evaluating integration and modularity in the input data, we recall on the Escoufier coefficient from morphometric studies on biological shapes, as a scalar measure of the co-variation between sets of landmarks. Statistical significance of the Escoufier coefficients among multiple sets of landmarks is established by means of a permutation test that is extended to 3D spatially dense landmark data. The spectral clustering described in this work results in finding the correct patterns in a more robust and accurate way compared to other unsupervised clustering techniques such as k-means or k-means++.

INTRODUCTION

All organisms are cohesive systems integrated to function as a whole (1,2). But instead of following uniform patterns of integration, many individuals display diverse and heterogeneous units due to their specific role (2). The human body is organized into multiple systems, each of them oriented to a specific task in accordance to its function, anatomy, and embryological origin. The facial shape represents a complex system that is experimentally and conceptually separable into several modules. For instance, bone and muscular cells for locomotion, skin cells for transpiration, protection, temperature regulation, and retinal cells for image processing – just to cite some trivial but effective examples. In the past 20 years of research in molecular and systems biology, the integration into subunits has been categorized under the heading of modularity (3–5), which implies the division of a structure into several parts that display high integration within the variables (modules), but few and weak interactions across the modules.

The fusion of modules into a compact structure can be studied by means of morphometric tools. The first challenge is defining the margins of each module and evaluating the hypotheses about these boundaries. Assumptions on the limits of each module can be studied by segmenting the structure in several ways and finding a suitable metric to evaluate the strength of the association between the sets of points. For morphometric data sets, hypotheses of modularity are commonly evaluated using the Escoufier's (1973) RV

coefficient (2), a ratio that represents the amount of covariation scaled by the amounts of variation within the two sets of variables.

We propose to investigate the 3D facial shape from a modular and thus local perspective. The translation of the concept of modularity from evolutionary and developmental biology to Image Analysis is possible through statistical shape modeling. In the context of facial morphology, a module represents a set of points that are similar with respect to a certain property, like the degree of correlation and their relative distance.

In this work, we deploy a normalized spectral clustering on a dataset of spatially-dense 3D landmark configuration of faces. The evaluation of the strength of covariation between the modules is done through the multiple RV coefficient (2) and its significance is tested by means of a permutation test extended to 3D sets of points.

MATERIALS AND METHODS

Population sample

592 Research participants from three West African/European admixed populations were selected for 3D facial images (6). They were collected in the United States (N=154, State College, PA, Williamsport, PA, and The Bronx, NY); Brasilia, Brazil (N=191); and Cape Verde (N=247, São Vicente, and Santiago), all under a Penn State University Internal Review Board (IRB) approved research protocol titled, "Genetics of Human Pigmentation, Ancestry

and Facial Features." Self-reported ancestry and sex were collected by survey. Ancestry informative markers (AIMs) were used to estimate individual genomic ancestry from DNA. 68-AIM ancestry estimates were generated using ADMIXMAP.

3D facial images

Three-dimensional images composed of surface and texture maps were taken using the 3dMDface system (3dMD, Atlanta, GA) (6). Participants were asked to hold their faces with a neutral expression and close their mouth for the picture. An anthropometric mask of 7,150 quasi-landmarks was non-rigidly mapped onto the original 3D images and their reflections (7,8). Subsequently, a Generalized Procrustes superimposition corrects for changes in position, orientation and scale of both the original and reflected configuration combined. After Procrustes superimposition, a single shape can be decomposed into its asymmetric and bilaterally symmetric component. The average of an original and its reflected configuration constitutes the symmetric component while the difference between the two configurations constitutes the asymmetric component. We work on facial shapes using the symmetric component only.

Face segmentation

The spectral clustering implementation used to divide the face into modules follows the work of Ng and Weiss (9). In general, given a set of data

points $x_i, \ldots, x_j$ and some notion of similarity $s_{ij} \geq 0$ between all pairs of data points x_i and x_j , the aim of clustering is to split the data points into groups such that points belonging to one group have a similar property and points in different groups are dissimilar to each other. The clustering problem can be formulated in terms of a similarity graph *G = (V, E)*, where each vertex v_i in the graph represents a data point x_i. Two vertices are connected if the similarity between the two is above a certain threshold and the edge between them is weighted by their similarity. The clustering problem then becomes a search for a partition of the graph such that the edges within a group are very high – meaning that data points are similar -, while edges between groups are very low – meaning that data points are very dissimilar. The crucial point of spectral clustering is the construction of the similarity graph, which models the local neighborhood relationships between data points. Our data-driven segmentation method makes use of the full set of faces. Given an average mesh configuration of the whole dataset, we first build the affinity matrix that encodes the similarity of each pair of landmarks. Assuming the data set $D \in \Re^{F\ x\ 3N}$ with *F* number participants and *N* 3D vertices, we split *D* into three subsets $D_{i=\{x,y,z\}} \in \Re^{F\ x\ N}$, each containing the corresponding spatial coordinates of the vertices. The correlation matrices $C_{i=\{x,y,z\}} \in \Re^{N\ x\ N}$ are computed from $D_{i=\{x,y,z\}} \in \Re^{F\ x\ N}$ and then averaged into the correlation matrix $\overline{C}$. Besides being grouped together according to their correlation value, points

close to each other on the facial surface should be clustered in the same module. Therefore, the spatial constrain is represented in the geodesic distance matrix $G \in \Re^{N\,x\,N}$, where each point on the matrix represents the closest path to reach a point (10). In order to combine the correlation contribution to the spatial component, the geodesic distance matrix needs to be normalized within the interval [0,1] using the Gaussian kernel $\overline{G} = e^{-G/2\sigma^2}$, where σ decides the decay of the Gaussian function and we take it equal to the mean of the geodesic distance matrix. The affinity matrix looks as follows (11):

$$A = \lambda * \overline{C} + (1 - \lambda) * \overline{G}, \qquad (1)$$

where λ determines the weight of the two components. The value $\lambda = 0.7$ represents a good choice to highlight the contribution of the inter-vertex correlation over the distances on the mesh.

Finally, the spectral clustering is performed as defined in (12). Briefly, the normalized Laplacian of the affinity matrix explained in equation (1) is computed and its top-k eigenvectors (corresponding to the k smallest eigenvalues) are retained. The clustering is performed through the traditional k-means algorithm on the first eigenvectors of the Laplacian matrix, initializing the centers with a weighted probability distribution in order to account for variability due to seed selection.

Statistical tests of covariation

Since the whole set of faces is segmented in k =1, …, 20 number of clusters, a multi-set measure of association is needed. The multiple Escoufier coefficient, RV_m, is the simple average of pairwise RV coefficients and represents a suitable statistic test of significance. The pairwise RV value is defined as follows:

$$RV_{ij} = \frac{trace(\Sigma_{ij} * \Sigma_{ji})}{\sqrt{trace(\Sigma_{ii} * \Sigma_{ii}) trace(\Sigma_{jj} * \Sigma_{jj})}}, \quad (2)$$

where Σ_{ij} is the covariance matrix of the two sets of variables i and j, and Σ_{ii} and Σ_{jj} are the variances within each set.

The multi-set RV coefficient is given by

$$RV_m = \frac{2}{k(k-1)} \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} RV_{ij}, \quad (3)$$

The statistical significance of the covariation between modules is usually established on 2D landmark data through permutation tests. In order to simulate the null hypothesis of complete independence between modules of 3D spatially dense landmarks, the clustering labels of each set are randomly reshuffled t=100 number of times so that any association between sets is due to chance only. At each t run, the RV_m coefficient is computed for the permuted modules and compared to the true value. The proportion of permutation rounds in which the RV_m of the permuted modules matches or exceeds the true RV_m value determines the significance level of the test.

The random regions at each t iteration are built by arbitrarily choosing k centers (according to the k number of clusters), and each k-th random module contains the closest landmarks based on the pairwise geodesic distance value.

RESULTS

Figure 1 shows the affinity matrix formulated in (1). It is an *NxN* (*N* number of landmarks) symmetric and square matrix with 1s along the diagonal, meaning that each landmark shows maximum association with itself. The range of values of the affinity matrix lies in the interval [0,1], where a value of 0 means that two points are dissimilar, while a value of 1 indicates they are both highly correlated and spatially close to each other.

Some examples of segmented faces are provided in Figure 2. It is striking how the first modules clearly separate the upper and lower face (k=2), and the area of the cheeks next to the ear is well defined (k=3). From k=4 we highlight that the nose, the mouth and the area of the cheeks next to the nose bridge are combined into a single compact module. It is worth noticing how the distinctiveness of several anatomic structures like the eyes, the mouth, the nose, the chin and the forehead start being independently identified for k=10 clusters.

The results of the permutation tests (t=100) in terms of multiple Escoufier coefficients and p-values are shown in Table 1. The multiple RV coefficient can be considered as an overall test of integration among the subsets, such that a low RV_M value indicates weak inter-cluster interactions, while a high value shows a strong association among modules. In Figure 3 we emphasize the decrease of the RV values with the number of modules. After a certain number of k modules there is a plateau with little significant variation.

Some examples of permuted modules are displayed in Figure 4, showing that each cluster is compact and of equal size to the correspondent true one.

DISCUSSION AND CONCLUSION

We have implemented a clustering algorithm to divide the face into modules, and have evaluated the strength of the association between them by means of statistical tools used in evolutionary developmental biology. The face is divided into modules according to its structural information only, namely the correlation between 3D spatially dense landmarks and their relative geometric location. The standard 2D statistical testing is expanded to deal with 3D mesh data. Therefore, the strength of covariation between sets is evaluated by means of a permutation test, where at each run a multi-set RV value is computed.

Clustering algorithms are widely used methods for exploratory data analysis, with several applications in statistics, biology, computer science or psychology. A common and often used technique is k-means (13), which seeks to minimize the average squared distance between points in the same cluster. It is a fast iterative algorithm, but lacks in finding the optimal cluster configuration.

An improvement of the kmeans algorithm is achieved with kmeans++ (14), which ameliorates the seed selection according to a weighted probability distribution proportional to the distance between centers.

Compared to the standard techniques such as kmeans or kmeans++, spectral clustering has many advantages (12): it outperforms the classical methods, it

has a simple implementation and can be solved by standard linear algebra methods. The added value of spectral clustering is that it finds partitions that are consistent with what the human eye would have chosen (9).

A motivating application of facial surface modularization is gender and ancestry prediction. Sex and ancestry are known in multimodal biometric systems as soft biometric traits because they give ancillary information used by humans to distinguish their peers, but lack the distinctiveness and permanence of some primary characteristics like the face, fingerprints, hand geometry or iris (15). The results in (16) confirm the hypothesis that investigating the 3D face for a *macro-to-micro* perspective enhances gender and ancestry predictions.

The challenging question concerns the number of clusters to use. The stopping criterion is application dependent, and we show that for sex and ancestry prediction (16) a suitable condition is taken according to the accuracy values of the prediction performance.

There are certainly some further improvements of our work. The affinity matrix built as in (1) does not take into account the whole spatial variability of each landmark, and it gives at the same time a strong constraint to spatially close landmarks – which should be avoided in order to explore the entire pairwise-landmark networks. Instead, we could compute the Escoufier coefficient between pairwise landmarks which represents the amount of covariation scaled by the amounts of variation within two landmarks. Some

added value we propose for spectral clustering could be its hierarchical implementation. First, the whole face could be divided into two modules through the normalized spectral clustering. Each module is then iteratively split into two, and the process continues. The idea of module propagation is intuitive since the landmarks that have been grouped separately at a specific number of clusters, should not be jointly modelled at a consequent split. Compared to standard clustering methods where at different number of clusters correspond different configurations, our hierarchical clustering would already retain the valuable information held in a previous segmentation step in order to build new sub modules.

AKNOWLEDGEMENTS

This work is supported by the Research Program of the Fund for Scientific Research – Flanders (Belgium) (FWO) and the Research Fun K.U. Leuven.

REFERENCES

1. Wagner GP, Pavlicev M, Cheverud JM. The road to modularity. Nat Rev Genet. 2007 Dec;8(12):921–31.

2. Klingenberg CP. Morphometric integration and modularity in configurations of landmarks: tools for evaluating a priori hypotheses. Evol Dev. 2009 Aug;11(4):405–21.

3. Wagner, Gunter P. W Gunter P. Homologues, natural kinds and the evolution of modularity. Am Zool. 1996;36(1):36–43.

4. Bolker JA. Modularity in Development and Why It Matters to Evo-Devo. Am Zool. 2000 Oct;40(5):770–6.

5. von Dassow G, Munro E. Modularity in animal development and evolution: elements of a conceptual framework for EvoDevo. J Exp Zool. 1999 Dec 15;285(4):307–25.

6. Claes P, Liberton DK, Daniels K, Rosana KM, Quillen EE, Pearson LN, et al. Modeling 3D Facial Shape from DNA. Luquetti D, editor. PLoS Genet. 2014 Mar 20;10(3):e1004224.

7. Claes P, Walters M, Clement J. Improved facial outcome assessment using a 3D anthropometric mask. Int J Oral Maxillofac Surg. 2012 Mar;41(3):324–30.

8. Claes P. A robust statistical surface registration framework using implicit function representations - application in craniofacial reconstruction. Proefschrift voorgedragen tot het behalen van het doctoraat in de ingenieurswetenschappen, K.U.Leuven, Leuven, Belgium.; 2007.

9. Yair Weiss N Andrew Y.Micheal I Jordan. On Spectral Clustering: Analysis and an algorithm. Adv Neural Inf Process Syst. 2002;2:849–56.

10. Tenenbaum JB, de Silva V, Langford JC. A global geometric framework for nonlinear dimensionality reduction. Science. 2000 Dec 22;290(5500):2319–23.

11. Tena JR, De la Torre F, Matthews I. Interactive region-based linear 3D face models. In ACM Press; 2011 [cited 2016 Apr 29]. p. 1. Available from: http://portal.acm.org/citation.cfm?doid=1964921.1964971

12. von Luxburg U. A tutorial on spectral clustering. Stat Comput. 2007 Dec;17(4):395–416.

13. MacQueen, J. B. Some Methods for classification and Analysis of Multivariate Observations. Proc 5th Berkeley Symp Math Stat Probab Univ Calif Press. 1967;pp. 281–97.

14. Arthur, David and Sergei Vassilvitskii. k-means++: The advantages of careful seeding. Proc Eighteenth Annu ACM-SIAM Symp Discrete Algorithms Soc Ind Appl Math. 2007;pp. 1027–36.

15. Xiaoguang Lu, Karthik Nandakumar and Unsang Park J Anil K. Integrating faces, fingerprints, and soft biometric traits for user recognition. ECCV Workshop BioAW. 2004;4.

16. Sero D., Shriver D. M., Vandermeulen D., Claes P. Modularity of 3D faces for gender and ancestry prediction. SHAPE Proc 2015.

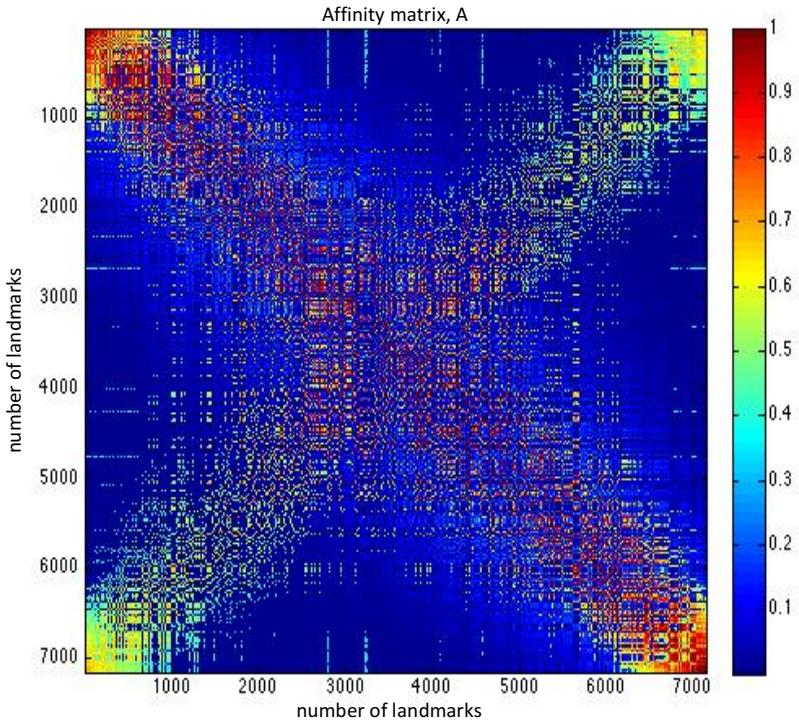

Figure 1. The image shows the affinity matrix build in (1). It is the result of the added contribution of the inter-vertex correlation and distance on the mesh. The impact of the pairwise-landmark correlation component is weighted more than the spatial component through the weighting factor λ. The dimension of the matrix corresponds to the number of dense 7,150 quasi-landmarks.

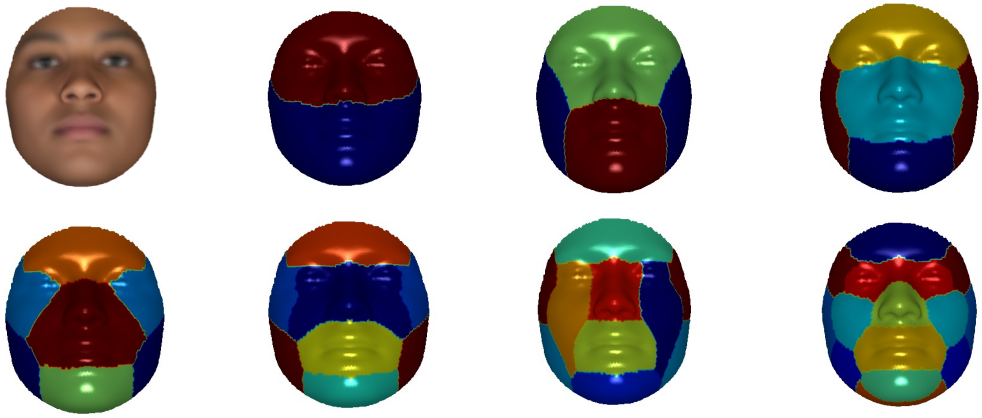

Figure 2. (from up left to bottom right). The average mesh configuration with texture; some examples of clustered faces with 2, 3, 4, 5, 6, 7 and 10 modules. It is remarkable that the first divisions clearly separate the upper and lower face (k=2), and the area of the cheeks next to the ear is well defined (k=3) and will be separate on further splits as well. From k=4 on, we highlight that the nose, the mouth and the area of the cheeks next to the nose bridge are combined into a single compact module. It is worth noticing how the distinctiveness of several anatomic structures like the eyes, the mouth, the nose, the chin and the forehead start being independently modelled for k=10 clusters.

modules	RV_M	p-value
2	0,6242	0,01
3	0,593	0,01
4	0,4921	0
5	0,4805	0
6	0,4171	0
7	0,3962	0
8	0,3823	0
9	0,4111	0
10	0,423	0
11	0,3994	0
12	0,3892	0
13	0,3719	0
14	0,3876	0
15	0,3829	0
16	0,383	0
17	0,3715	0
18	0,3462	0
19	0,3537	0
20	0,3497	0

Table 1. The table shows the values of the multiple Escoufier coefficient corresponding to each number of modules together with their p-value from the permutation test with t=100.

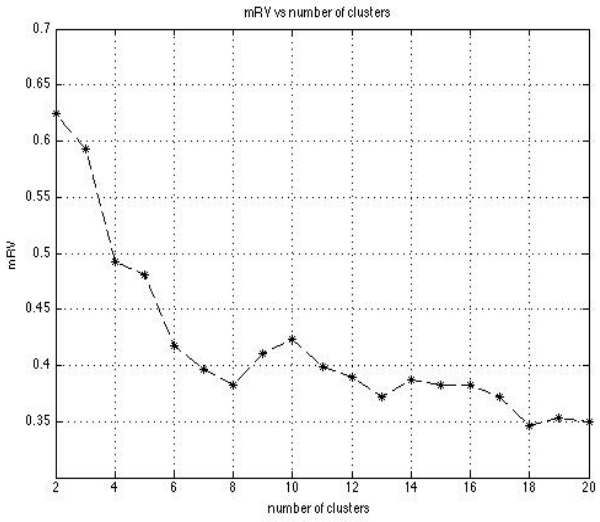

Figure 3. The plot of the RV_M behaviour shows a clear decrease with the number of clusters followed by a stagnation phase were little significant fluctuations.

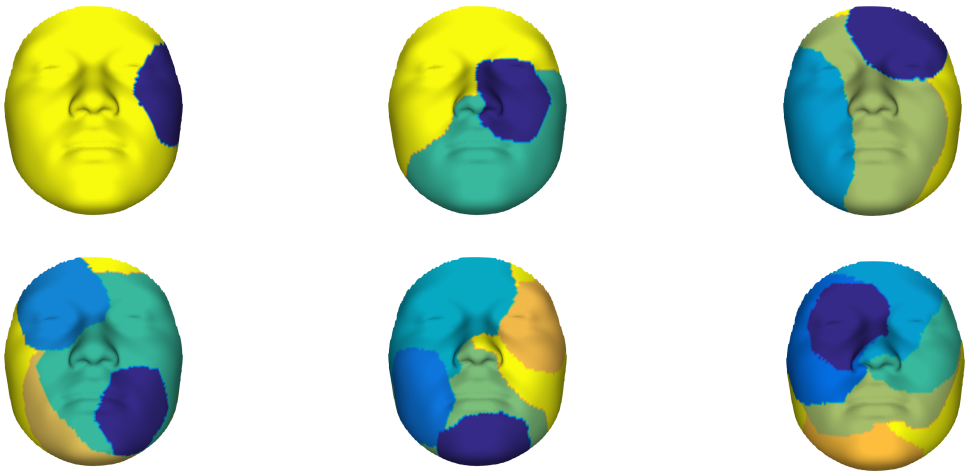

Figure 4. (from top left to bottom right) The random faces generated in the permutation tests using 2, 3, 4, 5, 6, and 7 modules. The number of landmarks in each module is equal to the original features in the true clusters.